UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/729,164	12/05/2003	Ajay Divakaran	MERL-1470	8566

22199 7590 09/14/2007

MITSUBISHI ELECTRIC RESEARCH LABORATORIES, INC.
201 BROADWAY
8TH FLOOR
CAMBRIDGE, MA 02139

EXAMINER	
SAINT CYR, LEONARD	

ART UNIT	PAPER NUMBER
2626	

MAIL DATE	DELIVERY MODE
09/14/2007	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary	Application No.	Applicant(s)	
	10/729,164	DIVAKARAN ET AL.	
	Examiner	Art Unit	
	Leonard Saint-Cyr	2626	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *10 July 2007*.

2a)☒ This action is **FINAL**. 2b)☐ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-13, 15 - 20* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-13, 15 -20* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____.
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

DETAILED ACTION

Response to Arguments

1. Applicant's arguments with respect to claims 1 -18 have been considered but are moot in view of the new ground(s) of rejection.

Response to Amendment

2. The amendment to the claims 19, and 20 filed on 07/10/07 does not comply with the requirements of 37 CFR 1.121(c) because a marked up version of the amended claims is not provided. Thus, the newly amended claims and claim 21 are not considered. Amendments to the claims filed on or after July 30, 2003 must comply with 37 CFR 1.121(c) which states:

(c) *Claims.* Amendments to a claim must be made by rewriting the entire claim with all changes (*e.g.*, additions and deletions) as indicated in this subsection, except when the claim is being canceled. Each amendment document that includes a change to an existing claim, cancellation of an existing claim or addition of a new claim, must include a complete listing of all claims ever presented, including the text of all pending and withdrawn claims, in the application. The claim listing, including the text of the claims, in the amendment document will serve to replace all prior versions of the claims, in the application. In the claim listing, the status of every claim must be indicated after its claim number by using one of the following identifiers in a parenthetical expression: (Original), (Currently amended), (Canceled), (Withdrawn), (Previously presented), (New), and (Not entered).

(1) *Claim listing.* All of the claims presented in a claim listing shall be presented in ascending numerical order. Consecutive claims having the same status of "canceled" or "not entered" may be aggregated into one statement (*e.g.*, Claims 1–5 (canceled)). The claim listing shall commence on a separate sheet of the amendment document and the sheet(s) that contain the text of any part of the claims shall not contain any other part of the amendment.

(2) *When claim text with markings is required.* All claims being currently amended in an amendment paper shall be presented in the claim listing, indicate a

status of "currently amended," and be submitted with markings to indicate the changes that have been made relative to the immediate prior version of the claims. The text of any added subject matter must be shown by underlining the added text. The text of any deleted matter must be shown by strike-through except that double brackets placed before and after the deleted characters may be used to show deletion of five or fewer consecutive characters. The text of any deleted subject matter must be shown by being placed within double brackets if strike-through cannot be easily perceived. Only claims having the status of "currently amended," or "withdrawn" if also being amended, shall include markings. If a withdrawn claim is currently amended, its status in the claim listing may be identified as "withdrawn—currently amended."

(3) *When claim text in clean version is required.* The text of all pending claims not being currently amended shall be presented in the claim listing in clean version, *i.e.*, without any markings in the presentation of text. The presentation of a clean version of any claim having the status of "original," "withdrawn" or "previously presented" will constitute an assertion that it has not been changed relative to the immediate prior version, except to omit markings that may have been present in the immediate prior version of the claims of the status of "withdrawn" or "previously presented." Any claim added by amendment must be indicated with the status of "new" and presented in clean version, *i.e.*, without any underlining.

(4) *When claim text shall not be presented; canceling a claim.*

(i) No claim text shall be presented for any claim in the claim listing with the status of "canceled" or "not entered."

(ii) Cancellation of a claim shall be effected by an instruction to cancel a particular claim number. Identifying the status of a claim in the claim listing as "canceled" will constitute an instruction to cancel the claim.

(5) *Reinstatement of previously canceled claim.* A claim which was previously canceled may be reinstated only by adding the claim as a "new" claim with a new claim number.

Claim Rejections - 35 USC § 103

3. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

4. Claims 1 – 5, 8, 11 –12, 14, 17, 19, and 20 are rejected under 35 U.S.C. 103(a) as being unpatentable over Leonardi et al., (Semantic Indexing if Multimedia

Documents, April –June 2002), in view of Barnard (Modelling and recognition of multi-modal temporal events, October 2002).

As per claim 1, Leonardi et al., teach a method for detecting highlights from videos, comprising:

extracting audio features from the video ("divide the input stream into audio and video"; page 46, col.2, lines 39 – 43);

classifying the audio features as labels (page 47, col.1, lines 9 – 14);

extracting visual features from the video ("divide the input stream into audio and video"; page 46, col.2, lines 39 – 43);

classifying the visual features as labels ("two-state-HMM classifier"; page 47, col.1, lines 38 – 43); and

fusing ("jointly consider audio and visual signals"), probabilistically("calculated four different performance indices"), the audio labels and visual labels to detect highlights in the video ("identifying relevant situations in soccer sequences"; page 49, col.1, lines 3 – 9; page 47, col.2, lines 1 – 7; page 44, col.2, lines 10, and 11).

However, Leonardi et al., do not specifically teach fusing into a single discrete-observation coupled hidden Markov model to detect highlight in the video.

Barnard teaches detecting particular event or actions within a television sport broadcast using the video and audio signals. A CHMM for accurately model interactions between speech and vision data streams (page 5, lines 3 – 6; page 1, lines 3 – 5).

Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to use a CHMM as taught by Barnard in Leonardi et al., because that would help better detect relevant actions within a television sport.

As per claim 2, Barnard further discloses that the video is compressed and the single discrete-observation coupled hidden Markov model includes the audio features, the visual features, audio states of the audio features and visual states of the visual features ("showing dependencies between the hidden states of the model"; figure 4; page 5, lines 3 – 8).

As per claim 3, Leonardi et al., further disclose that silent features are classified according to audio energy and zero cross rate ("extracts a feature vector from the low-level acoustic properties of each clip such as zero crossing rate"; page 46, col.2, lines 46 – 50).

As per claim 4, Leonardi et al., further disclose that the audio features are MeL-scale frequency cepstrum coefficients (page 46, col.2, lines 46 – 50).

As per claim 5, Leonardi et al., further disclose that the audio features are MPEG-7 descriptors (analyzed several samples from MPEG-7 using the proposed classification implies using MPEG-7 descriptors; page 50, col.2, line 13).

As per claim 8, Leonardi et al., further disclose the visual features are based on motion activity descriptors ("motion vectors" page 46, col.2, lines 50 – 53).

As per claim 11, Leonardi et al., further disclose the motion activity is averaged to obtain the visual labels (page 45, col.2, lines 36 – 38).

As per claim 12, Leonardi et al., further disclose the visual labels are selected from the group consisting of close shot, replay, and zoom (page 46, col.2, lines 1 – 12; col.2, lines 1 - 6).

As per claim 14, Barnard further discloses the discrete-observation coupled hidden Markov model includes audio hidden Markov models and visual hidden Markov models (fig.4; page 5, lines 3 – 8)).

As per claim 17, Leonardi et al., further disclose the video is a sport video ("soccer video"; page 45, col.1, lines 1 and 2).

As per claims 19, and 20, Leonardi et al., further disclose the audio portion of the video is compressed, and the visual portion of the video is compressed ("MPEG-7 content of audio-visual program"; page 50, col.2, lines 12 – 20).

5. Claims 6, and 7 are rejected under 35 U.S.C. 103(a) as being unpatentable over Leonardi et al., (Semantic Indexing if Multimedia Documents, April –June 2002), in view

of Barnard (Modelling and recognition of multi-modal temporal events, October 2002), and further in view Rui et al., (Automatically Extracting Highlights for TV Baseball Programs, Eighth ACM International Conference on Multimedia, pp.105 – 115, 2000)

As per claim 6, Leonardi et al., in view of Barnard do not specifically teach that the audio features are classified using Gaussian mixture models.

Rui et al., teach excited speech classification using Gaussian fitting (Gaussian fitting suggests Gaussian mixture; section 6.5, lines 1 and 2).

Therefore it would have been obvious to one of ordinary skill in the art at the time the invention was made to use Gaussian mixture models as taught by Rui et al., in Leonardi et al., in view of Barnard, because that would help better classify the audio signal.

As per claim 7, Leonardi et al., further disclose that audio labels are selected from the group consisting of applause, cheering, and music ("background noise" page 47, col.1, lines 12 – 14).

However, Leonardi et al., in view of Barnard do not specifically teach audio labels are selected from the group consisting of ball hit, speech with music, male speech and female speech.

Rui et al., teach classifying audio signals into silence, speech, music, song, and mixtures of the above, Baseball hit detection (section 5.2; section 2, paragraph 6, lines 11, and 12).

Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to classify the audio signals as taught by Rui et al., in Leonardi et al., in view of Barnard, because that would help better determine the highlights of the soccer video.

The examiner takes official notice that classifying speech between male speech and female speech is well known in the art. One having ordinary skill in the art would have found it obvious to classify the audio as male speech and female speech, because that would help determine particular scenes of the multimedia documents.

6. Claims 9, 10, and 15 are rejected under 35 U.S.C. 103(a) as being unpatentable over Leonardi et al., (Semantic Indexing if Multimedia Documents, April –June 2002), in view of Barnard (Modelling and recognition of multi-modal temporal events, October 2002), and further in view Wang et al., (Integration of Multimodal Features For Video Scene Classification based on HMM, 1/99).

As per claim 9, Leonardi et al., further disclose that visual features include motion vectors ("motion vectors" page 46, col.2, lines 50 – 53).

However, Leonardi et al., in view of Barnard do not specifically teach that visual features include dominant color.

Wang et al., teach visual features include the most dominant color (page 54, paragraph 2, lines 6, and 7).

Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to include the most dominant color in visual features as

taught by Wang et al., in Leonardi et al., in view of Barnard, because that would help better classify the video signal, so that highlights can be found.

As per claim 10, Leonardi et al., in view of Barnard do not specifically teach that the variance of the motion activity is quantized to obtain the visual labels.

Wang et al., teach that visual features include the most dominant color, the most dominant notion vectors, and the mean and variance of motion vector. We quantize the colors of each video frame into 64 colors adaptively (page 54, paragraph 2, lines 6 – 9).

Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to quantize the variance as taught by Wang et al., in Leonardi et al., in view of Barnard, because that would help better classify the video signal, so that highlights can be found.

As per claim 15, Leonardi et al., in view of Barnard do not specifically teach that the discrete-observation coupled hidden Markov model is generated from a Cartesian product of states of the audio hidden Markov models and the visual hidden Markov models, and a Cartesian product of observations of the audio hidden Markov models and the visual hidden Markov models.

Wang et al., teach training an HMM for each of the audio, color, and motion modalities separately. The observed sequences of different features are fed into the corresponding HMM. The final observation probability is computed as... (page 55, paragraph 2).

Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to calculate Cartesian product of HMMs as taught by Wang et al., in Leonardi et al., in view of Barnard, because that would help determine particular scenes of the multimedia documents.

7. Claims 16, and 18 are rejected under 35 U.S.C. 103(a) as being unpatentable over Leonardi et al., (Semantic Indexing if Multimedia Documents, April –June 2002), in view of Barnard (Modelling and recognition of multi-modal temporal events, October 2002), and further in view of Rui et al., (US PAP 2003/0103647).

As per claim 16, Leonardi et al., training the discrete-observation coupled hidden Markov model ("training two-state HMM"; page 47, col.1, lines 40, and 41).

However Leonardi et al., in view of Barnard do not specifically teach training with hand labeled videos.

Rui et al., teach that training set is view –labeled in that each face image is manually labeled (paragraph 95).

Therefore it would have been obvious to one of ordinary skill in the art at the time the invention was made to manually label videos as taught by Rui et al., in Leonardi et al., in view of Barnard, because that would help better classify the video signals.

As per claim 18, Leonardi et al., in view of Barnard do not specifically teach determining likelihoods for the highlights; and thresholding the highlights.

Rui et al., disclose that multi-cue tracking module includes an observation likelihood module (paragraph 109); detecting candidates for new face regions, wherein each candidate is a region of the video content that potentially includes a new face. Generating a confidence level for each candidate, if the confidence level does not exceed the threshold value, the candidate is discarded (paragraph 41).

Therefore it would have been obvious to one of ordinary skill in the art at the time the invention was made to threshold candidate face regions as taught by Rui et al., in Leonardi et al., in view of Barnard, because that would help determine particular scenes by rejecting non relevant scenes.

Conclusion

8. **THIS ACTION IS MADE FINAL.** Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the mailing date of this final action.

9. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Leonard Saint-Cyr whose telephone number is (571) 272-4247. The examiner can normally be reached on Mon- Friday.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Richemond Dorvil can be reached on (571) 272-7602. The fax phone number for the organization where this application or proceeding is assigned is (571)-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

LS
09/06/07

RICHEMOND DORVIL
SUPERVISORY PATENT EXAMINER